CLAIMS

1.   A process for identifying statistically-outlying data points in at least one dataset, comprising:

    a)   receiving the at least one dataset; and

    b)   identifying the statistically-outlying data points present in the at least one dataset based on the information contained in the at least one dataset.

2.   The process of claim 1, wherein the at least one dataset comprises data associated with levels of gene expression obtained under two different conditions.

3.   The process of claim 2, wherein the two different conditions reflect an occurrence of at least one of a physiological process, a pathophysiological process, an oncogenic process, a mutational process, a pharmacologically-induced process, an immuno-precipitation-induced process, and a developmental process.

4.   The process of claim 1, further comprising one or more of the following steps:

    c)   storing the at least one dataset in a matrix;

    d)   shifting each row of the matrix by a center of mass of the at least one dataset;

    e)   computing a principal axis of the at least one dataset;

    f)   rotating the at least one dataset so that the principal axis coincides with x-axis; and

    g)   generating strip functions that define boundaries outside which the statistically-outlying data points in the at least one dataset are located.

5.   The process of claim 4, wherein the at least one dataset comprises the set $E = \{\chi_i\}_{i=1}^N$ of N points in R^D.

6.   The process of claim 4, wherein the strip functions that define boundaries that identify the statistically-outlying data points present in the at least one dataset are generated by computing a stopping point F_Q using a top-down procedure.

5

7.   The process of claim 6, wherein the strip functions are smoothed by averaging of the strips generated from more than one determination.

8.   The process of claim 6, wherein a stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $F_{Q'} > \alpha_0$.

9.   The process of claim 6, wherein a stopping point in the computation of

10

F_Q is set at $Q' \in D(Q_0))$ if $\left| \tilde{Q} \right| < n_0$.

10.   The process of claim 6, wherein a stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\beta_{\tilde{Q}} > \delta_0$.

11.   The process of claim 6, wherein a stopping point in the computation of

F_Q is set at $Q' \in D(Q_0))$ if $\left| \hat{Q}' \backslash \tilde{Q} \right| > \alpha_1 \cdot \left| \tilde{Q}' \right|$.

15

12.   The process of claim 6, wherein the stopping point in the computation of F_Q is applied twice.

13.   A software arrangement operable by a processing arrangement for identifying the statistically-outlying data points present in at least one dataset based on the information contained in the at least one dataset, the software arrangement comprising:

20

a)   a first set of instructions operable to configure the processing arrangement to receive the at least one dataset; and

b)   a second set of instructions operable to configure the processing arrangement to identify the statistically-outlying data points

-23-

present in the at least one dataset based on the information contained in the at least one dataset.

14.   The software arrangement of claim 13, wherein the at least one dataset comprises data associated with levels of gene expression obtained under two different conditions.

15.   The software arrangement of Claim 14, wherein the two different conditions reflect an occurrence of at least one of a physiological process, a pathophysiological process, an oncogenic process, a mutational process, a pharmacologically-induced process, an immuno-precipitation-induced process, and a developmental process.

16.   The software arrangement of claim 13, further comprising at least one of the instructions:

c)   a third set of instructions operable to configure the processing arrangement to store the at least one dataset in a matrix;

d)   a fourth set of instructions operable to configure the processing arrangement to shift each row of the matrix by a center of mass of the at least one dataset;

e)   a fifth set of instructions operable to configure the processing arrangement to compute a principal axis of the at least one dataset;

f)   a sixth set of instructions operable to configure the processing arrangement to rotate the at least one dataset so that the principal axis coincides with x-axis; and

g)   a seventh set of instructions operable to configure the processing arrangement to generate strip functions that define boundaries outside which the statistically-outlying data points in the at least one dataset are located.

17.   The software arrangement of claim 16, wherein the at least one dataset comprises a set $E = \{\chi_i\}_{i=1}^{N}$ of N points in R^D.

18.   The software arrangement of claim 16, wherein the strip functions that define boundaries that identify the statistically-outlying data points present in the at least one dataset are generated by computing a stopping point F_Q using a top-down procedure.

19.   The software arrangement of claim 18, wherein the strip functions are smoothed by averaging of the strips generated from more than one determination.

20.   The software arrangement of claim 18, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $F_{Q'} > \alpha_0$.

21.   The software arrangement of claim 18, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\left|\tilde{Q}\right| < n_0$.

22.   The software arrangement of claim 18, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\beta_{\tilde{Q}} > \delta_0$.

23.   The software arrangement of claim 18, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\left|\hat{Q}' \backslash \tilde{Q}\right| > \alpha_1 \left|\tilde{Q}'\right|$.

24.   The software arrangement of claim 18, wherein the stopping point in the computation of F_Q is applied twice.

25.   A storage medium which includes thereon a software arrangement to be executed by a processing arrangement for identifying the statistically-outlying data points present in the at least one dataset based on the information contained in the at least one dataset, the software arrangement comprising:

a)   a first set of instructions operable to configure the processing arrangement to receive the at least one dataset; and

b)   a second set of instructions operable to configure the processing arrangement to identify the statistically-outlying data points present in the at least one dataset based on the information contained in the at least one dataset.

5

26.   The storage medium of claim 25, wherein the at least one dataset comprises data associated with levels of gene expression obtained under two different conditions.

10   27.   The storage medium of claim 26, wherein the two different conditions reflect the occurrence of at least one of a physiological process, a pathophysiological process, an oncogenic process, a mutational process, a pharmacologically-induced process, an immuno-precipitation-induced process, and a developmental process.

15   28.   The storage medium of claim 25, wherein the software arrangement further comprises at least one of the following instructions:

c)   a third set of instructions operable to configure the processing arrangement to store the at least one dataset in a matrix;

d)   a fourth set of instructions operable to configure the processing
20   arrangement to shift each row of the matrix by a center of mass of the at least one dataset;

e)   a fifth set of instructions operable to configure the processing arrangement to compute a principal axis of the at least one dataset;

25   f)   a sixth set of instructions operable to configure the processing arrangement to rotate the at least one dataset so that the principal axis coincides with x-axis; and

g)   a seventh set of instructions operable to configure the processing arrangement to generate strip functions that define

boundaries outside which the statistically-outlying data points in the at least one dataset are located.

29. The storage medium of claim 28, wherein the at least one dataset comprises a set $E = \{\chi_i\}_{i=1}^N$ of N points in R^D.

30. The storage medium of claim 28, wherein the strip functions that define boundaries that identify the statistically-outlying data points present in the at least one dataset are generated by computing at stopping point F_Q using a top-down procedure.

31. The storage medium of claim 30, wherein the strip functions are smoothed by the averaging of the strips generated from more than one determination.

32. The storage medium of claim 30, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $F_{Q'} > \alpha_0$.

33. The storage medium of claim 30, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\left|\tilde{Q}\right| < n_0$.

34. The storage medium of claim 30, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\beta_{\tilde{Q}} > \delta_0$.

35. The storage medium of claim 30, wherein the stopping point in the computation of F_Q is set at $Q' \in D(Q_0))$ if $\left|\hat{Q}' \setminus \tilde{Q}\right| > \alpha_1 \left|\tilde{Q}'\right|$.

36. The storage medium of claim 30, wherein the stopping point in the computation of F_Q is applied twice.

37. A system comprising:

a processing arrangement operably configured to:

a)   receiving the at least one dataset; and

b)   identifying the statistically-outlying data points present in the at
        least one dataset based on the information contained in the at
        least one dataset.

38.   The system of claim 37, further comprising further processing
        arrangement configured to generate the at least one dataset.

39.   The system of claim 38, further comprising a detector configured to
        detect a plurality of signals indicative of gene expression and convert
        the detected signals into the at least one dataset.

40.   The method of claim 1, wherein the at least one data set comprises data
        associated with financial trends.

41.   The software arrangement of claim 13, wherein the at least one data set
        comprises data associated with financial trends.

42.   The storage medium of claim 25, wherein the at least one data set
        comprises data associated with financial trends.